External Validation Guidance and Toolkit: Method 3 Supplement

Table of Contents

External Validation Facility Selection Methods Comparison			
	Method 1 - Prioritizing Facilities with Highest Likelihood of Event Occurrence	**Method 2 - CAD Approach**	**Method 3- Stratified Random Sample**
Target criteria	This method prioritizes facility selection based on highest likelihood of event occurrence. It is more likely to select facilities with higher patient volume, and thus a higher predicted/expected number of events.	This method prioritizes facility selection by difference between predicted and observed number of events (CAD). It focuses on facilities with negative CAD values. These facilities reported zero or very few events and have a high predicted number of events.	This method prioritizes selecting facilities randomly to create representative estimates.
What type of facilities are selected?	Focuses on larger healthcare facilities with high exposure volume, and thus high predicted/expected events.	Focuses on potential under reporters: facilities that reported very few events yet have a high predicted number of events.	Focuses on reviewing a representative sample of the state/jurisdiction
Ranking algorithm	Facility ranking algorithm uses predicted events and facility standardized infection ratio (SIR) values for ranking and selection. SIR is a ratio of observed vs. predicted events and is subject to variability. A small facility with low predicted volume of events with even one observed event could have a high SIR value.	Facility ranking algorithm uses CAD. CAD metric is robust, stable and reflects the true facility HAI burden.	No facility ranking algorithm used. Facilities are stratified by bed size.
Which method should my agency use?	Agencies with no prior validation history should use Method 1 to determine HAI misclassification patterns. If external agencies are already aware of underreporting concerns, they may select Method 2.	Agencies with previous validation history that identified underreporting as a potential concern should use Method 2.	Agencies who wish to evaluate a representative sample of their jurisdiction.
Number of facilities	• 20 or fewer facilities: validate them all • 21 to 149 facilities: at least 18 targeted facilities plus a 5% random sample of remaining facilities • 150 or more facilities: select at least 21 targeted facilities plus a 5% random sample of remaining facilities.	Fewer than 30 facilities: validate them all 30 or more facilities: 30 facilities, distributed between Stratum 1 and 2	Fewer than 30 facilities: validate them all 30 or more facilities: 30 facilities, distributed between Stratum 1 and 2
Charts per facility	60	40	40

Method 3 Overview

This Method 3 document is a supplement to the NHSN 2022 Toolkit and Guidance for External Validation (2022 External Validation Toolkit). The instructions that follow provide a third potential approach for selection of facilities for validation and instructions for selection of medical records

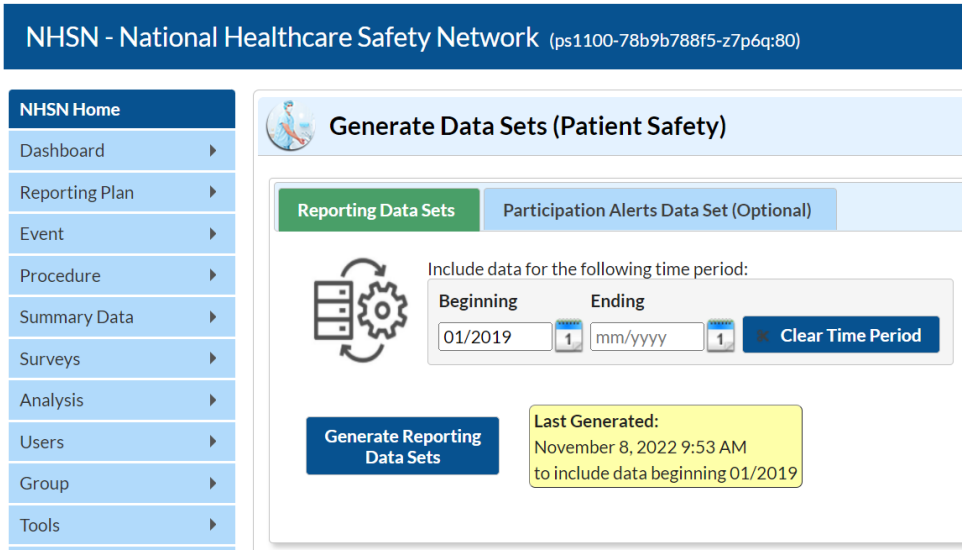
Stratified Random Sampling

This is the third method auditors may choose from for facility and medical record selection. If HAI's or other events will be validated in multiple facility types, separate sampling should be completed for acute care hospitals (ACH), long-term acute care hospitals (LTACH), and inpatient rehabilitation facilities (IRF). This will provide a system for stratified random sampling with the aim of producing a representative sample to measure inter-rater reliability of the measure(s) in the jurisdiction undergoing review.

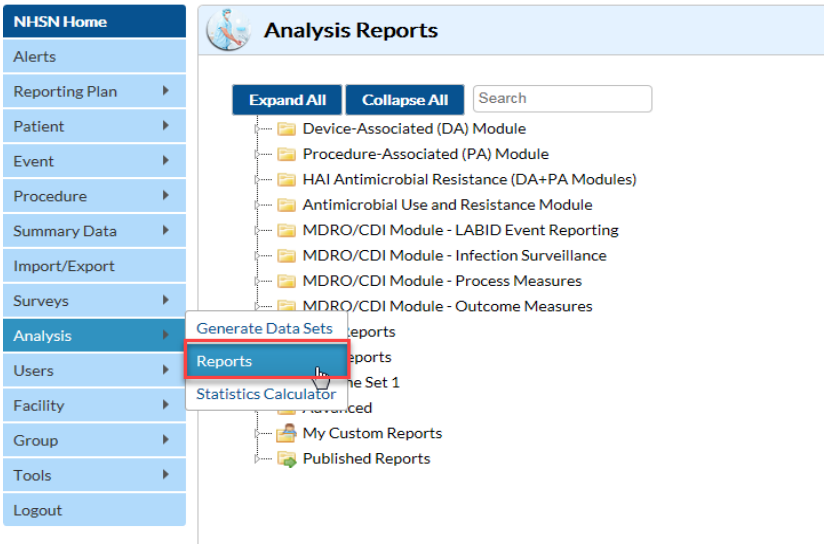
Create an Acute Care Hospital Facilities List

Generate datasets and obtain a list of all facilities reporting the HAI to use in creating a random sample:

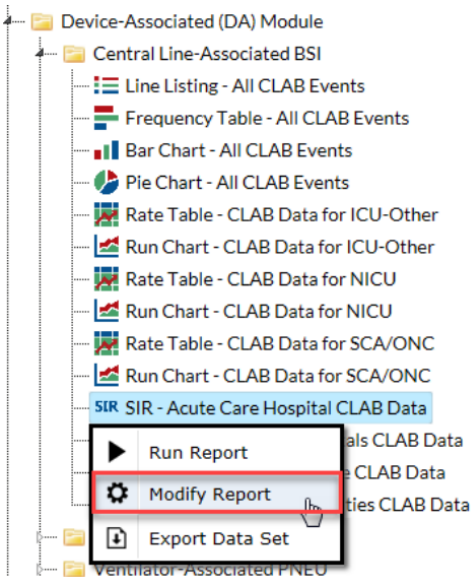
1. Generate new datasets in NHSN to ensure any data updates are included for analysis. On the NHSN Landing Page, navigate to Patient Safety Component --> [YOUR State Users' Group]. Select the "Analysis" tab and click "Generate Data Sets." Click the Generate New button. Allow the dataset generation process to complete; you can leave NHSN during the generation process.



2. After successful dataset generation, navigate to Analysis -> Reports to display the tree view list of all analysis reports available within NHSN's analysis tool.

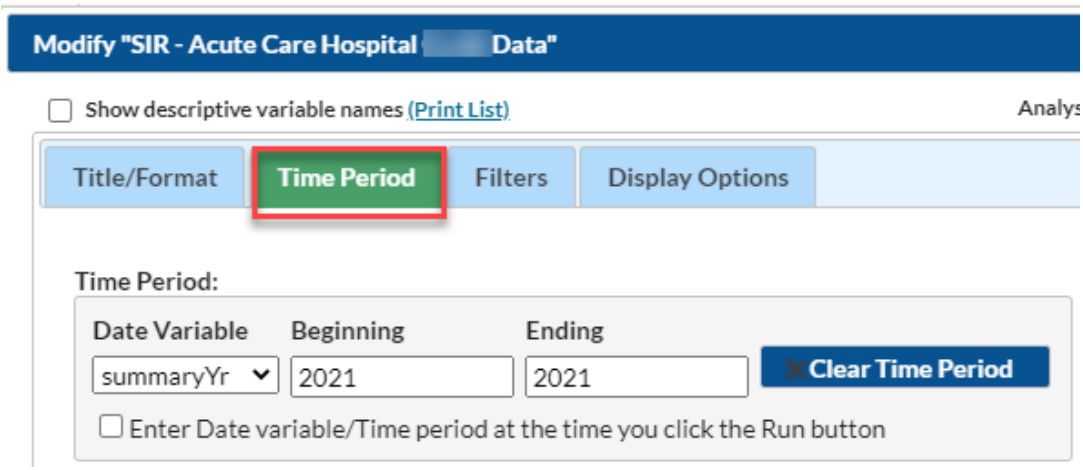


3. Use the tree view structure to navigate to the SIR report of interest for the HAI selected for validation. In this example for validation of CLABSI, select the Device Associated Module -> Central Line-Associated BSI -> SIR Acute Care Hospital CLAB Data. The report uses data reported to NHSN that has been shared with the group. Click the Modify button to proceed to the modification screen, which can be used to filter and export data from NHSN.

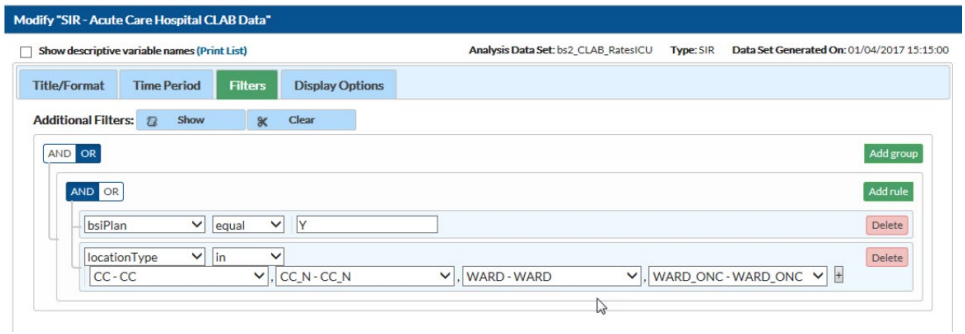


4. A modification screen will open with two key areas to modify: one that controls the time interval of data that are analyzed and displayed and one that controls the level of aggregation of that data.

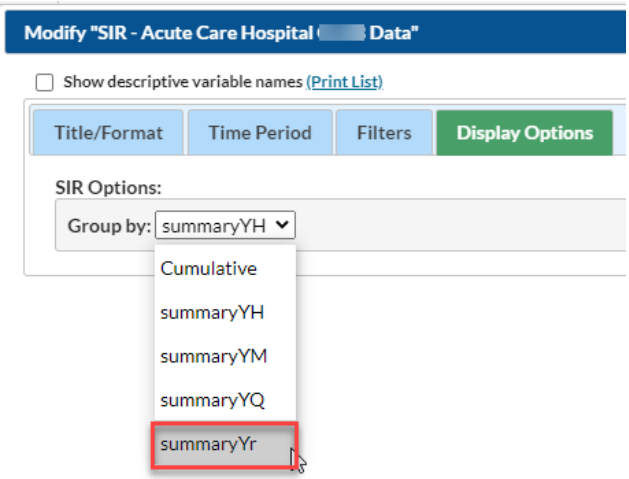
   4a. Use the “Time Period” tab to limit the time period of data that is included in the report to be exported. Set “Date Variable” to SummaryYr, “Beginning” and “Ending” to whatever year you will be validating. This example is validating 2021 data:



4b. Navigate to the "Filters" tab. The parameters needed will vary by HAI. For this example, focusing on CLABSI, select bsiPlan from the dropdown list, set the next field to "equal" and the next as "Y". Add another rule by selecting "Location type" from the dropdown list and set the next field to "in" and Value(s) to "CC-CC" and "CC_N-CC_N" to specify all ICU locations, adult and neonatal. Add any additional validation locations desired following the same process. Scroll to the bottom of the pop-up screen and select "Save". For filter modifications on all other HAIs, see section 3.3 of the 2022 External Validation Toolkit



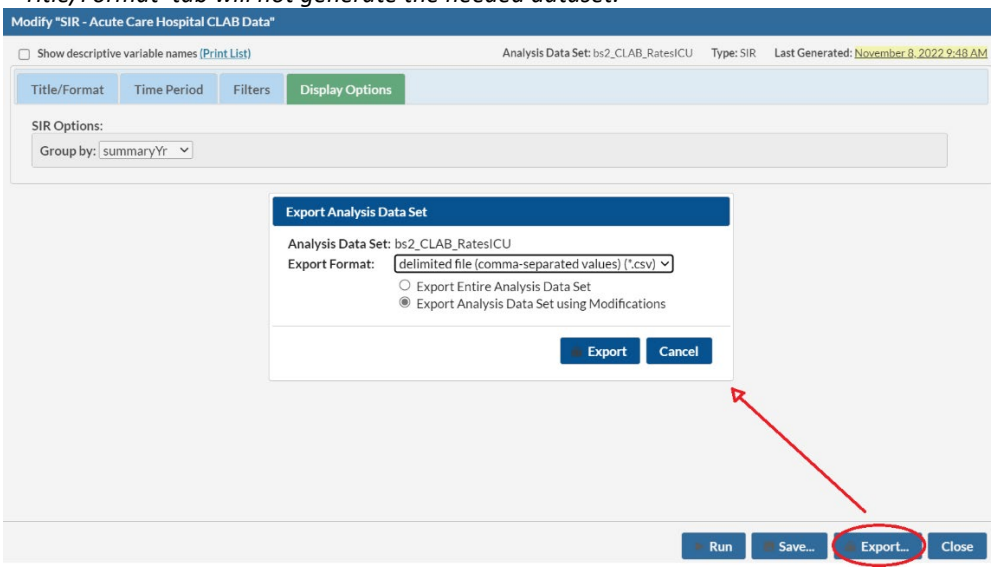
4c. Under the "Display Options" section, use the "Group by" option to view the data at a particular level of aggregation. Change the Group by option to "SummaryYr".



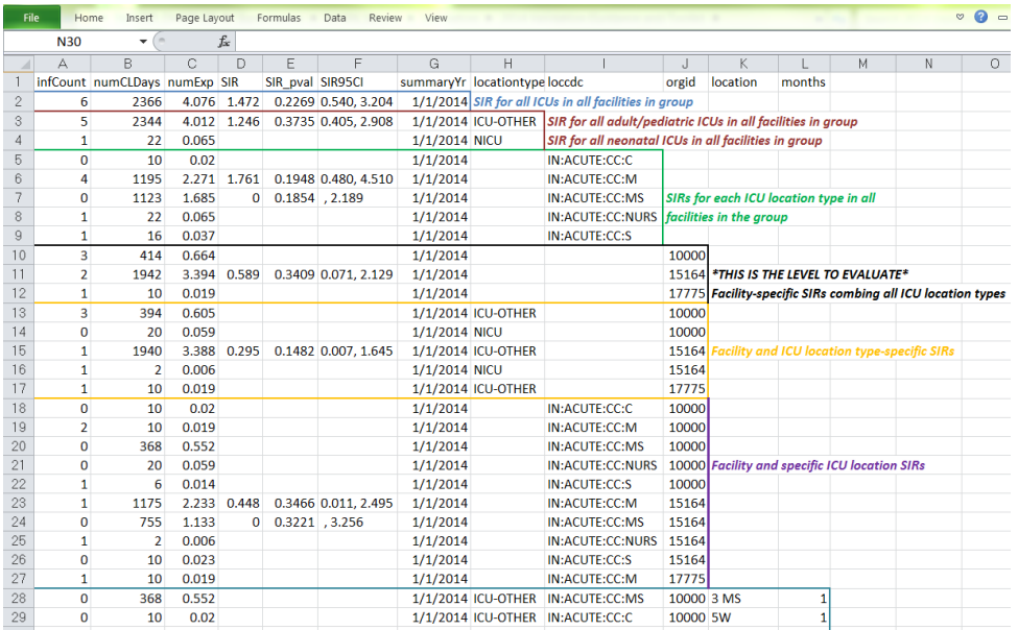
5. After making these modifications, click on the Export button at the bottom of the screen (not the Run button). This will prompt the "Export Analysis Data Set" dialog box to appear. Use the dropdown

menu to select the file format to export the data. In this example, we will export a comma-separated value (*.csv) dataset. Make sure the bottom radio button is selected and click the Export button to begin the export process. NHSN will create a .zip file with your data export in it.

Note: *You must use the export button to generate these results. Selecting an excel file from the 'Title/Format' tab will not generate the needed dataset.*



6. In Excel, you will need to find the level of aggregation at the "orgID" level, so you have an unduplicated list of all facilities reporting data for the HAI during the timeframe you specified in step 4a. An easy way to find this level of aggregation is to look at the "loccdc" column and scroll until it is blank. The unique orgIDs will then begin to list in ascending order. Once you see the list go through the highest orgID and start over at the smallest orgID, that is where the unduplicated facility list ends. You will also notice that once the list of orgIDs starts over, the column "locationtype" will begin to have data as well. In this example, the desired aggregation level starts at row 10, ends at row 12, and is outlined in black. *Tip: the columns "loccdc" and "locationtype" are* ***blank*** *for the rows you want.*



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	infCount	numCLDays	numExp	SIR	SIR_pval	SIR95CI	summaryYr	locationtype	loccdc	orgid	location	months			
2	6	2366	4.076	1.472	0.2269	0.540, 3.204	1/1/2014								
3	5	2344	4.012	1.246	0.3735	0.405, 2.908	1/1/2014	ICU-OTHER							
4	1	22	0.065				1/1/2014	NICU							
5	0	10	0.02				1/1/2014		IN:ACUTE:CC:C						
6	4	1195	2.271	1.761	0.1948	0.480, 4.510	1/1/2014		IN:ACUTE:CC:M						
7	0	1123	1.685	0	0.1854	, 2.189	1/1/2014		IN:ACUTE:CC:MS						
8	1	22	0.065				1/1/2014		IN:ACUTE:CC:NURS						
9	1	16	0.037				1/1/2014		IN:ACUTE:CC:S						
10	3	414	0.664				1/1/2014			10000					
11	2	1942	3.394	0.589	0.3409	0.071, 2.129	1/1/2014			15164					
12	1	10	0.019				1/1/2014			17775					
13	3	394	0.605				1/1/2014	ICU-OTHER		10000					
14	0	20	0.059				1/1/2014	NICU		10000					
15	1	1940	3.388	0.295	0.1482	0.007, 1.645	1/1/2014	ICU-OTHER		15164					
16	1	2	0.006				1/1/2014	NICU		15164					
17	1	10	0.019				1/1/2014	ICU-OTHER		17775					
18	0	10	0.02				1/1/2014		IN:ACUTE:CC:C	10000					
19	2	10	0.019				1/1/2014		IN:ACUTE:CC:M	10000					
20	0	368	0.552				1/1/2014		IN:ACUTE:CC:MS	10000					
21	0	20	0.059				1/1/2014		IN:ACUTE:CC:NURS	10000					
22	1	6	0.014				1/1/2014		IN:ACUTE:CC:S	10000					
23	1	1175	2.233	0.448	0.3466	0.011, 2.495	1/1/2014		IN:ACUTE:CC:M	15164					
24	0	755	1.133	0	0.3221	, 3.256	1/1/2014		IN:ACUTE:CC:MS	15164					
25	1	2	0.006				1/1/2014		IN:ACUTE:CC:NURS	15164					
26	0	10	0.023				1/1/2014		IN:ACUTE:CC:S	15164					
27	1	10	0.019				1/1/2014		IN:ACUTE:CC:M	17775					
28	0	368	0.552				1/1/2014	ICU-OTHER	IN:ACUTE:CC:MS	10000	3 MS	1			
29	0	10	0.02				1/1/2014	ICU-OTHER	IN:ACUTE:CC:C	10000	5W	1			

7. Once you identify where the aggregation at orgID starts, click on the first orgID cell and drag until you reach the highest value (before it starts to repeat). Copy the selected cells and paste into a new Excel worksheet or a new sheet within the same worksheet. This is your final list of all unduplicated facilities reporting the HAI of interest during the timeframe you specified. You will use this list to generate your random sample.

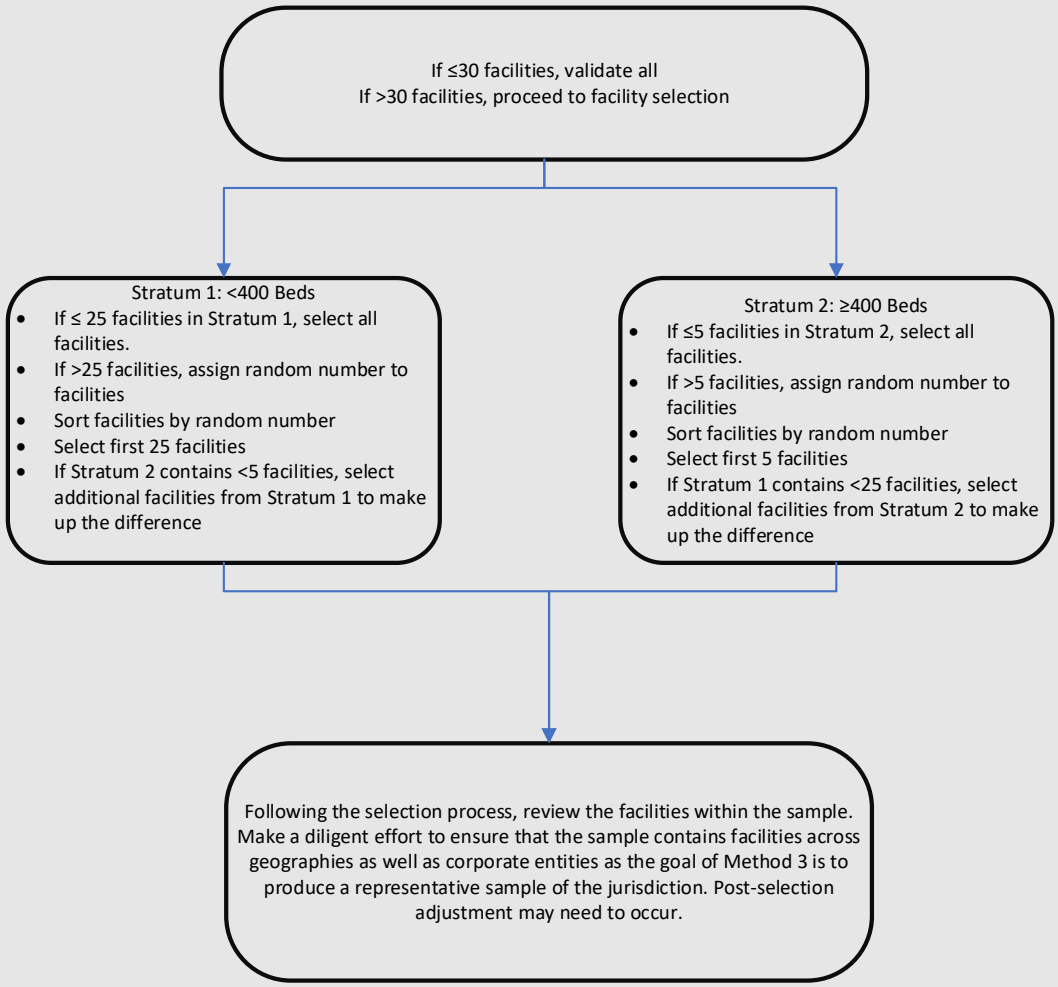
Determine the Minimum Facility Sample Size

Use the facility list generated from NHSN to determine the appropriate minimum sample size. This approach to ACH facility selection is designed to prioritize measurement of a representative sample of facilities to produce an estimate of the inter-rater reliability of the HAI event determination. A recommended minimum number of facilities should be validated (with a recommended minimum number of medical records) for each selected HAI:

- Smaller states/jurisdictions with 30 or fewer facilities should validate all facilities
- Larger states with more than 30 facilities should select at least 30 facilities to review based on the selection criteria below.

Algorithm for Selecting Acute Care Hospital Facilities

Selection of facilities: Use these instructions if the sampling frame consists of greater than 30 facilities



1. Divide the total facilities in the sampling frame into two strata:
    a. Stratum 1: Includes all facilities in the sampling frame that have a bed size of <400
    b. Stratum 2: Includes all facilities in the sampling frame that have a bed size of ≥400.
2. Stratum 1:
    a. If there are 25 or fewer facilities within Stratum 1, select all facilities within Stratum 1.
        i. Proceed to Stratum 2.
    b. If there are more than 25 facilities within Stratum 1, assign a random number to each facility.
        i. Refer to Table 1 below for two methods for random number assignment
    c. Select the first 25 facilities from the randomized facility list
3. Stratum 2:
    a. If there are 5 or fewer facilities within Stratum 2, select all facilities within Stratum 2.
        i. Return to Stratum 1. Continuing in descending order, select additional facilities for a total of 30 facilities selected.
    b. If there are more than 5 facilities within Stratum 2, assign a random number to each facility.
    c. Select the first 5 facilities from the randomized facility list.
    d. If Stratum 1 had fewer than 25 facilities, return to Stratum 2.
    e. Continuing in descending order, select additional facilities for a total of 30 facilities selected

Table 1. Random number assignment methods

Option 1: Excel	1. Using the facility list created above, or an HAI linelist, insert the command =ROUND(RAND()*1000000,0) into column B and drag to paste this command for each row of the facility list. This will generate a random number for each orgID. 2. Select and copy the values from column B and use the Paste Special (Paste Values) feature to paste the number values into column C. Note: any edit made to the Excel sheet will cause the numbers in column B to recalculate. This is normal and can be ignored if you have an iteration copied. 3. Delete column B so the columns shift left and column C becomes column B. 4. Sort by column B, making sure column A is included in the sort (click on "Expand selection" if a dialog box appears). This is your final list that has been assigned and sorted by a random number.
Option 2: Random Number Generator Website + Excel	1. Identify the total number of facilities from the list created above, or the number of records on HAI line list, 2. Go to https://www.random.org/sequences/ 3. Input 1 as the smallest value, and the total number of facilities/records as the largest value, and click "Get Sequence" 4. Copy the sequence created and paste it into column B of your spreadsheet. 5. Sort by column B, making sure column A is included in the sort (click on "Expand selection" if a dialog box appears). This is your final list that has been assigned and sorted by a random number.
Option 3: SAS Codes	1. Enter the appropriate file path where prompted in the code 2. For medical record random number generation, determine if you need/want the program to create an 'EoC' number. If yes, run code as written. If no, delete the lines of code as specified in the program, then run code. 3. The final list, assigned and sorted by a random number, will be exported to the same folder specified in step 1.

Medical Record Selection

For sampling, a medical record refers to the record of an inpatient or observation admission to a single facility, also referred to as an episode of care. For surgical procedures, the episode of care refers to the procedure and all associated medical encounters documented during the surveillance follow-up window.
For each HAI to be validated, a sample size of 40 Medical Records/Episodes of Care per facility is recommended as a goal. The approach provided for the selection of medical records produces a sampling frame designed to target records that could have events.

1. Before requesting medical records or other data for the audit, download ("freeze") the facilities reported data from NHSN. See Chapter 2.7 in the 2022 External Validation Toolkit for detailed step by step instructions for each HAI.
2. Request facilities to send line lists of candidate HAI cases from the validation locations for the validation timeframe. These line lists will include positive blood and urine cultures, COLO and HYST procedures, positive CDI results and positive MRSA blood specimen results. Facilities should be encouraged to provide the line lists in an Excel template (Refer to Appendix 1.1).
3. Use the medical record selection process below for the HAI of interest
4. For Medical Record Abstraction Templates (MRAT) and instructions, please see the HAI-specific MRAT guidance documents located here: 2022 PSC Data Validation Resources | NHSN | CDC

Medical Record Selection Process:

For CLABSI, CAUTI, COLO and HYST validation a sampling frame of eligible medical records will be developed for each HAI and from these 40 episodes of care will be selected, by targeting those with increased risk of event occurrence. For MRSA bacteremia and CDI LabID Event validation, enrichment criteria are based on a positive laboratory test.
Definitions of targeted populations for each type of HAI and enrichment criteria used to select those with increased likelihood of HAI identification are described below. A total of 40 episodes of care will be reviewed for each HAI per facility.

Sample structure

- 40 medical records each for CLABSI in validation locations, CAUTI in validation locations, COLO, HYST, MRSA bacteremia LabID Events, and CDI LabID Events:
    - For CLABSI in validation locations, these will be stratified by NICU and adult/pediatric ICU locations, other validation locations, and will prioritize targeted pathogens.
    - For COLO and HYST, the medical record at the time of the surgical procedure will be reviewed, as well as any additional records during the 30-day surveillance period.
    - For MRSA bacteremia LabID Events and CDI LabID Events, 40 episodes of care per HAI will be reviewed from targeted positive lab findings.

A. CLABSI in validation locations Targeted Medical Record Selection Process

1. From each selected facility, request a securely transmitted line listing of all positive blood cultures (PBC), from all validation locations reporting to NHSN, for the entire validation timeframe, with required additional variables used for medical record identification and matching to NHSN reports (See Appendix 1.1 for recommended line listing structure and template facility letters). Facilities should be instructed to provide this in a spreadsheet (e.g. Excel) format.

2. Assure the line list includes PBCs from all validation locations required to report CLABSIs to NHSN, using location mapping information in NHSN for reference.
3. Assign a random number, using the steps outlined in Table 1, to each PBC.
4. Sort the list of PBCs by MRN and admission date to generate clusters of blood specimens with same MRN and within the same admission date, also called episodes of care. Create an 'episode of care' (EoC) field where the first EoC = 1. All PBCs in the episode should have the same EoC number.
5. Review each PBC and assign to stratum 1 if identified organism is on "Targeted Pathogen" list (see list below) or assign the PBC to stratum 2 if organism is not on the targeted pathogen list.
6. Use location information to identify if PBC was associated with NICU vs. adult/pediatric ICU records and create a NICU (Yes/No) field. Assign NICU status to each PBC as appropriate.
    a) If facility has no NICU, skip to step 8 below, and select 10 additional medical records from adult/pediatric ICUs for screening sample.
7. Select the NICU screening sample:
    a) Select where NICU= Yes, and stratum = 1 (targeted pathogens)
    b) Sort by random number
    c) Select the first 10 random numbers with unique EoC numbers as the sample of NICU records containing candidate CLABSIs.
    d) If 10 NICU records with stratum 1 blood specimens are not available, supplement the NICU sample with NICU records with stratum 2 blood specimens (where NICU = Yes, and stratum = 2); select the initial records (lowest random numbers with unique EoC) to total 10 selected medical records from NICU.
8. Select the non-NICU screening sample
    a) Select where NICU = No, and stratum = 1 (targeted pathogens)
    b) Sort by random number
    c) Select the first 30 random numbers with unique EoC numbers as the sample of validation location records with candidate CLABSIs.
    d) If 30 validation location records with stratum 1 blood specimens are not available, supplement the non-NICU record sample with stratum 2 blood specimen (where NICU= No, and stratum = 2); select the initial records (lowest random numbers with unique EoC) to total 30 selected medical records from validation locations.
9. Review both samples to ensure no EoC are repeated/duplicated. If there are any EoC that are duplicated, keep the first sampled PBC and replace the subsequent samples from the pertinent screening sample lists.
10. The final screening sample should contain (up to) 40 medical records divided among NICU (if available) and other validation locations.
11. If medical records are not well balanced among different targeted pathogens, consider post-selection adjustment to include a variety of these organisms.

Targeted Pathogens:

1) *Candida spp*. (yeast)
2) *Enterococcus spp*.
3) *Staphylococcus aureus* (includes MRSA, MSSA)
4) Coagulase-negative staphylococcus (includes most staphylococcus spp. other than *S. aureus*, MRSA, MSSA)
5) *Klebsiella spp*., *E. coli*, or *Pseudomonas spp*. (common gram negatives)

Before requesting medical records for the audit, download ("freeze") the facility's reported data from NHSN

B. CAUTI IN VALIDATION LOCATIONS Medical Record Selection Process

1. From each selected facility, request a securely transmitted line listing of all positive urine cultures (PUC; Positive urine cultures with no more than 2 identified pathogens, with at least one bacterium with greater than or equal to 10^5 CFU/ml organisms), from all validation locations reporting to NHSN, for the entire validation timeframe, with required additional variables used for medical record identification and matching to NHSN reports (See Appendix 1.1 for recommended line listing structure). Facilities should be instructed to provide this in a spreadsheet (e.g. Excel) format
2. Assure the line list includes appropriate PUCs from all validation locations required to report CAUTIs to NHSN, using location mapping information in NHSN for reference.
3. Assign a random number to each PUC, following the steps outlined in Table 1.
4. Sort the list of urine cultures by MRN and admission date to generate clusters of urine cultures associated with same MRN and admission date, also called episodes of care. Create an 'episode of care' (EoC) field where the first EoC = 1 for the associated urine cultures. All PUCs in the episode should have the same EoC number.
5. Select simple random sample of (up to) 40 records in validation locations for review.
    a) Sort by random number
    b) Select the first 40 random numbers with unique EoC numbers.
    c) If there are any duplicate EoC numbers, keep the record with the smallest random number and substitute out the others with the next PUC on the list.
6. The final screening sample should contain: (up to) 40 medical records from validation locations.

Before requesting medical records for the audit, download ("freeze") the facility's reported data from NHSN

C. COLO Procedure Targeted Medical Record Selection Process

1. Using NHSN, download a line listing of all COLO procedures for the validation timeframe, following these steps (if not validating for 2021 data, update the year accordingly for the timeframe being validated):
    a) Log In to NHSN for the facility being validated and the Patient Safety Module.
    b) From the left-hand Navigation Bar, Click "Analysis" then "Reports."
    c) Select the folder titled "Advanced," then "Procedure-level Data," "Modify."
    d) Select the "Line Listing – All Procedures," Then the Modify Report button.
    e) Change the Title to "Line Listing – COLO Procedures 2021,"
    f) Under "Title/Format" Select Excel (xls), and consider whether you want to check the box for "Show Descriptive Labels." This option will make the variable names longer (and more explicit), but is often not necessary if you know the variable names.
    g) Select "time period"," then select ProcDateYr, for "Beginning" enter 2021, and for "Ending" enter 2021.
    h) Select Filter do the following:
        i. Under "add rule" use the drop down list and select "procCode"
        ii. In the next drop down list select equals , and the next "COLO-Colon surgery"
        iii. Select "add rule" to add another line. Using drop down list, select "outpatient", then equals "No"
        iv. select "ageAtProc"
        v. "ageAtProc", equals or greater "18"
    i) Select "Display Variables select "Modify List"; retain the default Selected Variables: orgID, patID, DOB, gender, procID, procDate, and procCode. Add variables by double clicking from the left-hand list: ProcDateYr, outpatient, ageAtProc (to assure that you have selected 2021 inpatient adult COLO procedures), anesthesia, asa, procDurationHr, procDurationMin, Scope, medAff,

numBeds, swClass, and bs2_modelRiskAdultAll (variable that will be used to select procedures at higher risk to result in SSI).

j) Select "Sort Variables; remove procCode from the right hand list by double clicking (all procedures will be COLO). Add procID by double clicking the variable in the left hand box; it will move to the right hand box. Click Save for future use if you wish: "Line Listing for COLO Procedures 2021."

k) Select Run. You should see a line listing in Excel sorted by procID from lowest to highest.

2. Select the 40 procedures with the highest SSI risk ("bs2_modelRiskAdultAll") for review.

Before requesting medical records for the audit, download ("freeze") the facility's reported data from NHSN

D. HYST Procedure Targeted Medical Record Selection Process

1. Using NHSN, download a line listing of all HYST procedures for the entire validation timeframe, following the steps outlined above for COLO.
    a. If you have saved your template for downloading the line list of COLO procedures, you can make a few small modifications to download the HYST procedures rather than starting over (where you have entered "COLO" replace it with "HYST).
2. Select the 40 procedures with the highest SSI risk ("bs2_modelRiskAdultAll") for review.

Before requesting medical records for the audit, download ("freeze") the facility's reported data from NHSN

E. Strategy for Selection of MRSA Bacteremia LabID Events for Validation

1. From each selected facility, request a securely transmitted line listing of all positive MRSA blood cultures from all inpatient locations/ED/24 hour observations for the validation timeframe, with required additional variables used for medical record identification and possible matching to NHSN reports (See Appendix 1.1 for recommended line listing structure).
2. Assign a random number to each positive MRSA blood culture in the list, following steps outlined in Table 1.
3. Sort the list by MRN, admission date, and specimen date to generate clusters of positive MRSA blood cultures with the same MRN and admission dates, also called episodes of care. Create an 'episode of care' (EoC) field where the first EoC = 1. All positive MRSA blood cultures in the episode should have the same EoC number.
4. Select simple random sample of (up to) 40 records in validation locations for review:
    a. Sort by random number
    b. Select the first 40 random numbers with unique EoC numbers
    c. If there are any duplicate EoC numbers, keep the record with the smallest random number and substitute out the others with the next positive MRSA blood culture on the list.
5. The final screening sample should contain: (up to) 40 medical records from validation locations.

Before requesting medical records or other data for the audit, download ("freeze") the facility's reported data from NHSN

F. Strategy for Selection of *C. difficile* Infection (CDI) LabID Events for Validation

1. From each selected facility, request a securely transmitted line listing of all positive *Clostridium difficile* stool specimens from all inpatient locations/ED/24 hour observations for the entire validation timeframe, with required additional variables used for medical record identification and possible matching to NHSN reports (See Appendix 1.1 for recommended line listing structure). Facilities should be instructed to provide this in a spreadsheet (e.g. Excel) format.
2. Assign a random number to each positive CDI result in the list, following the steps outlined in Table 1.
3. Sort the list by MRN, admission date, and specimen date to generate clusters of positive CDI result with the same MRN and admission dates, also called episodes of care. Create an 'episode of care' (EoC) field where the first EoC = 1. All positive CDI results in the episode should have the same EoC number.
4. Select simple random sample of (up to) 40 records in validation locations for review:
    a. Sort by random number
    b. Select the first 40 random numbers with unique EoC numbers
    c. If there are any duplicate EoC numbers, keep the record with the smallest random number and substitute out the others with the next positive CDI result on the list.
5. The final screening sample should contain: (up to) 40 medical records from validation locations.

Before requesting medical records or other data for the audit, download ("freeze") the facility's reported data from NHSN

Note: For ease of use and printing, Medical Records Abstraction Tools (MRATs) and Instructions for use of the MRATs are located in separate documents under Supporting Documents on the Validation webpage